

961414632

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

Color Space Coding Framework

Inventor(s):
Jacky Shen
Feng Wu
Lujun Yuan
Shipeng Li

ATTORNEY'S DOCKET NO. MS1-1806US

TECHNICAL FIELD

This invention relates to multimedia, and in particular to a color space coding framework for handling video formats.

BACKGROUND

The consumer electronics market is constantly changing. One reason that the market is constantly changing is that consumers are demanding higher video quality in their electronic devices. As a result, manufacturers are designing higher resolution video devices. In order to support the higher resolution video devices, better video formats are being designed that provide better visual quality.

There are two main color spaces from which the majority of video formats are derived. The first color space is commonly referred to as the RGB (Red Green Blue) color space (hereinafter referred to as RGB). RGB is used in computer monitors, cameras, scanners, and the like. The RGB color space has a number of formats associated with it. Each format includes a value representative of the Red, Green, and Blue chrominance for each pixel. In one format, each value is an eight bit byte. Therefore, each pixel consumes 24 bits (8 bits (R) + 8 bits (G) + 8 bits (B)). In another format, each value is 10 bits. Therefore, each pixel consumes 30 bits.

Another color space has been widely used in television systems and is commonly referred to as the YCbCr color space or YUV color space (hereinafter referred to as YUV). In many respects, YUV provides superior video quality in comparison with RGB at a given bandwidth because YUV takes into consideration that the human eye is more sensitive to variations in the intensity of a pixel than in its color variation. As a result, the color difference signal can be sub-sampled to achieve bandwidth saving. Thus, the video formats associated with the YUV color

space, each have a luminance value (Y) for each pixel and may share a color value (represented by U and V) between two or more pixels. The value of U (Cb) represents the blue chrominance difference between B-Y and the value of V (Cr) represents the red chrominance difference between R-Y. A value for the green chrominance may be derived from the Y, U, and V values. YUV color space has been used overwhelmingly in video coding field.

There are several YUV formats currently existing. FIGURES 1-5 illustrate five of the more common YUV formats: YUV444, YUV422, YUV420, YUV411, and YUV410, respectively. FIGURES 1-5 graphically illustrate arrays 100-500, respectively. The illustrated arrays are each eight by eight array of blocks. However, the arrays may be of any dimension and do not necessarily need to be square. Each block in the array (denoted by a dot) represents an array of pixels. For convenience and keeping with conventional video techniques, the following discussion describes each block as representing one pixel (e.g., pixels P_1 - P_4). Therefore, hereinafter, the term pixel will be used interchangeably with the term block when referring to arrays 100-500. The pixels are grouped into macroblocks (e.g., macroblocks MB_1 - MB_N) based on the sampling that is desired for the target video format. FIGURES 1-3 illustrate each macroblock having four pixels (e.g., P_1 - P_4). FIGURES 4-5 illustrate each macroblock having sixteen pixels (e.g., P_1 - P_{16}). Each of the YUV formats will now be described in more detail.

FIGURE 1 graphically illustrates the YUV444 format. In the YUV444 format, each pixel is represented by a Y, U, and V value. For example, for pixel P_1 , the YUV444 format includes eight bits for the Y_1 value, eight bits for the U_1 value, and eight bits for the V_1 value. Thus, each pixel is represented by twenty-four bits. Because this format consumes twenty-four bits for each pixel, other

1 YUV formats are down-sampled from the YUV444 format so that the number of
2 bits per pixel is reduced. The reduction in bits per pixel provides improvement in
3 streaming efficiency. However, down-sampling results in a corresponding
4 degradation in video quality.

5 FIGURE 2 graphically illustrates the YUV422 format. In the YUV422
6 format, each pixel is represented by a Y value. However, in contrast with the
7 YUV444 format, the U and V values are optionally filtered and then down-
8 sampled. The filtering and down-sampling may be performed simultaneously
9 using known techniques. Array 200 conceptually illustrates the results from the
10 down-sampling by illustrating every second horizontal pixel in the array 200 as
11 sampled. The sampled pixels are denoted with an "X" in array 200. Thus, pixels
12 P_1 and P_3 are each represented by twenty-four bits. However, pixels P_2 and P_4 are
13 each represented by eight bits (Y value only). The average number of bits per
14 pixel in the YUV422 format is sixteen bits $((24+24+8+8)/4)$. The YUV422 is a
15 packed YUV color space, which means that the Y, U, and V samples are
16 interleaved. Typically, standards that support the YUV422 format, such as MPEG-
17 2 and MPEG-4, code all the chrominance blocks together. For example, the
18 YUV422 format for MPEG-2 stores the YUV422 data in memory as Y1 U1 Y2
19 V1, where Y1 and Y2 represent the luminance value for pixels P_1 and P_2 ,
20 respectively. Y1 and Y2 represent two luminance blocks. U1 and V1 represent
21 two chrominance blocks.

22 FIGURE 3 graphically illustrates the YUV420 format. Array 300
23 conceptually illustrates the results from the optional filtering and down-sampling
24 from the YUV444 format by illustrating every second horizontal and every second
25 vertical pixel in the array 300 as sampled. Again, the sampled pixels are denoted

1 with an "X" in array 300. Thus, for the YUV420 format only pixel P_1 is
2 represented by twenty-four bits. Pixels P_2 - P_4 is each represented by eight bits (Y
3 value only). The average number of bits per pixel in the YUV420 format is twelve
4 bits $((24+8+8+8)/4)$. The YUV420 is a planar format, not a packed format. Thus,
5 the YUV420 data is stored in memory such that all of the Y data is stored first,
6 then the U data, then all of the V data. Therefore, there are four luminance blocks,
7 one U chrominance block and one V chrominance block.

8 FIGURE 4 graphically illustrates the YUV411 format. Array 400
9 conceptually illustrates the results from the optional filtering and down-sampling
10 from the YUV444 format by illustrating every fourth horizontal pixel in array 400
11 as sampled. Thus, pixels P_1 , P_5 , P_9 , and P_{13} are each represented by twenty-four
12 bits and the other twelve pixels are represented by eight bits. The average number
13 of bits per pixel in the YUV411 format is twelve bits.

14 FIGURE 5 graphically illustrates the YUV410 format. Array 500
15 conceptually illustrates the results from the optional filtering and down-sampling
16 from the YUV444 format by illustrating every fourth horizontal pixel and every
17 fourth vertical pixel in array 500 as sampled. Thus, only pixel P_1 is represented by
18 twenty-four bits and the other fifteen pixels are represented by eight bits. The
19 average number of bits per pixel in the YUV410 format is 10 bits.

20 Thus, based on the quality that is desired and the transmission bandwidths
21 that are available, an electronic device manufacturer may design their electronic
22 devices to operate with any of these and other formats. However, later when
23 transmission bandwidths increase and/or consumers begin to demand higher
24 quality video, the existing electronic devices will not support the higher quality
25 video format. For example, currently many digital televisions, set-top boxes, and

1 other devices are designed to operate with the YUV420 video format. In order to
2 please the different categories of consumers, there is a need to accommodate both
3 video formats.

4 Television stations could broadcast both the higher video format (e.g.,
5 YUV422) and the lower video format (e.g., YUV420). However, this option is
6 expensive to the television broadcasters because it involves having the same
7 content on two different channels, which consumes valuable channel resources.
8 Thus, currently, the higher resolution format is transcoded to the lower resolution
9 format either at the server side or at the client side. FIGURE 6 is a block diagram
10 illustrating the transcoding process. A transcoder 600 accepts an input format,
11 such as Format A (e.g., YUV422), and outputs an output format, such as Format B
12 (e.g., YUV420). During the transcoding process, the entire video input format is
13 decoded, which includes the Y, U, and V components. The Y component must be
14 decoded along with the UV components because the UV components are motion
15 compensated and the resultant motion vectors can only be obtained by decoding
16 the Y component. Thus, the luminance blocks and all the chrominance blocks are
17 decoded to get a reconstructed version of the original video in the input format.
18 Then, chrominance components are down-sampled to convert the input format to
19 the desired output format. Finally, the newly generated video is encoded again to
20 generate a bit stream in the output format (Format B). This transcoding process is
21 expensive because it is generally equivalent to an encoder plus a decoder. Fast
22 transcoding methods exist, but generally result in quality loss.

23 The transcoder 600 may exist at the client side, the server side, or at another
24 location. If the transcoding process is performed at the client side, consumers that
25 subscribe to the high quality video may access the high quality video while other

1 consumers can access the lower quality video. If the transcoding process is
2 performed at the server, none of the consumers can access the high quality video.
3 Neither option is optimal because the transcoding process is very expensive and
4 generally leads to quality degradation. Therefore, there is a need for a better
5 solution for providing high quality video while maintaining operation with
6 existing lower quality video devices.

7 **SUMMARY**

8 The present color space coding framework provides conversions between
9 one or more video formats without the use of a transcoder. A video information
10 stream that includes color information formatted in accordance with a first color
11 space sampling format is split into a base stream and an enhanced stream. The
12 base stream is formatted in accordance with a second color space sampling format.
13 The enhanced stream includes enhanced information that when combined with the
14 base stream re-constructs the first format. During encoding, the enhanced stream
15 may be encoded using spatial information related to the base information stream.
16 An output stream of the encoded base stream and encoded enhanced stream may
17 be interleaved, concatenated, or may include independent files for the encoded
18 base stream and the encoded enhanced stream.

19 20 **BRIEF DESCRIPTION OF THE DRAWINGS**

21 FIGURES 1-5 are a series of graphical depictions of various encoding
22 formats derived from the YUV color space.

23 FIGURE 6 is a block diagram of a transcoder for converting between two
24 different video formats.

1 FIGURE 7 illustrates an exemplary computing device that may utilize the
2 present exemplary coding framework.

3 FIGURE 8 is a block diagram of a chroma separator for separating a first
4 video encoded format into multiple streams in accordance with the exemplary
5 color space coding framework.

6 FIGURE 9 is a block diagram of a chroma compositor for merging the
7 multiple streams into the first video encoded format in accordance with the
8 exemplary color space coding framework.

9 FIGURE 10 is a graphical depiction of the first video encoded format and
10 the multiple streams after the chrominance blocks have been separated from the
11 first video encoded format by the chroma separator shown in FIGURE 8.

12 FIGURE 11 is a block diagram of an encoder which incorporates the
13 present color space coding framework.

14 FIGURE 12 is a block diagram of a decoder which incorporates the present
15 color space coding framework.

16 FIGURE 13 is a graphical representation of an exemplary bit stream for
17 transmitting the multiple bit streams shown in FIGURES 11 and 12.

18 FIGURE 14 is a graphical representation of another exemplary bit stream
19 for transmitting the multiple bit streams shown in FIGURES 11 and 12.

20 FIGURES 15 – 20 illustrate exemplary integer lifting structures suitable for
21 use in conjunction with FIGURES 8 and 9.

DETAILED DESCRIPTION

Briefly stated, the present color space coding framework provides a method for creating multiple streams of data from an input video encoded format. The multiple streams of data includes a base stream that corresponds to a second video encoded format and at least one enhanced stream that contains enhanced information obtained from the input video encoded format. By utilizing the present method, multimedia systems may overcome the need to transcode the input video format into other video formats in order to support various electronic devices. After reading the following description, one will appreciate that using the present color space coding framework, an electronic device configured to operate using a lower quality format may easily discard periodic chrominance blocks and still have the resulting video displayed correctly. The following discussion uses the YUV422 and YUV420 video formats to describe the present coding framework. However, one skilled in the art of video encoding will appreciate that the present coding framework may operate with other video formats and with other multimedia formats that can be separated into blocks with information similar to the information contained within the chromo blocks for video formats.

Thus, the following description sets forth a specific exemplary coding framework. Other exemplary coding frameworks may include features of this specific embodiment and/or other features, which aim to eliminate the need for transcoding multimedia formats (e.g., video formats) and aim to provide multiple multimedia formats to electronic devices.

The following detailed description is divided into several sections. A first section describes an exemplary computing device which incorporates aspects of the present coding framework. A second section describes individual elements

1 within the coding framework. A third section describes the exemplary bit streams
2 that are encoded and decoded in accordance with the present color space coding
3 framework.

4 **Exemplary Computing Device**

5 FIGURE 7 illustrates an exemplary computing device that may utilize the
6 present exemplary coding framework. An example of a computing device
7 includes a set-top box that enables a television set to become a user interface to the
8 Internet and enables the television set to receive and decode digital television
9 (DTV) broadcasts. In another configuration, the exemplary computing device may
10 be separate from the set-top box and provide input to the set-top box. Another
11 example of a computing device includes a video recording device, such as a digital
12 camcorder or digital camera. In a very basic configuration, computing device 700
13 typically includes at least one processing unit 702 and system memory 704.
14 Depending on the exact configuration and type of computing device, system
15 memory 704 may be volatile (such as RAM), non-volatile (such as ROM, flash
16 memory, etc.) or some combination of the two. System memory 704 typically
17 includes an operating system 705, one or more program modules 706, and may
18 include program data 707. A Web browser may be included within the operating
19 system 705 or be one of the program modules 706. The Web browser allows the
20 computing device to communicate via the Internet.

21 Computing device 700 may have additional features or functionality. For
22 example, computing device 700 may also include additional data storage devices
23 (removable and/or non-removable) such as, for example, magnetic disks, optical
24 disks, or tape. Such additional storage is illustrated in FIGURE 7 by removable
25

1 storage 709 and non-removable storage 710. Computer storage media may
2 include volatile and nonvolatile, removable and non-removable media
3 implemented in any method or technology for storage of information, such as
4 computer readable instructions, data structures, program modules, or other data.
5 System memory 704, removable storage 709 and non-removable storage 710 are
6 all examples of computer storage media. Computer storage media includes, but is
7 not limited to, RAM, ROM, EEPROM, flash memory or other memory
8 technology, CD-ROM, digital versatile disks (DVD) or other optical storage,
9 magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage
10 devices, or any other medium which can be used to store the desired information
11 and which can be accessed by computing device 700. Any such computer storage
12 media may be part of device 700. Computing device 700 may also have input
13 device(s) 712 such as keyboard, mouse, pen, voice input device, touch input
14 device, etc. Output device(s) 714 such as a display, speakers, printer, etc. may
15 also be included. These devices are well known in the art and need not be
16 discussed at length here. Computing device 700 may also have one or more
17 devices (e.g., chips) for video and audio decoding and for processing performed in
18 accordance with the present coding framework.

19 Computing device 700 may also contain communication connections 716
20 that allow the device to communicate with other computing devices 718, such as
21 over a network. Communication connections 716 are one example of
22 communication media. Communication media may typically be embodied by
23 computer readable instructions, data structures, program modules, or other data in
24 a modulated data signal, such as a carrier wave or other transport mechanism, and
25

1 includes any information delivery media. The term “modulated data signal”
2 means a signal that has one or more of its characteristics set or changed in such a
3 manner as to encode information in the signal. By way of example, and not
4 limitation, communication media includes wired media such as a wired network or
5 direct-wired connection, and wireless media such as acoustic, RF, infrared and
6 other wireless media. Thus, communication media includes telephone lines and
7 cable. The term computer readable media as used herein includes both storage
8 media and communication media.

9 Exemplary Coding Framework

10 FIGURE 8 is a block diagram of a chroma separator 800 for separating a
11 first video encoded format (e.g., Format A) into multiple streams (e.g., base format
12 B stream and enhanced format B stream). The process for separating the base
13 stream from format A is now described. Those skilled in the art will appreciate the
14 common practice of performing low pass filtering before down-sampling from a
15 higher resolution to a lower resolution in order to improve the quality of the down-
16 sampled format. Thus, the chroma separator 800 may include an optional low pass
17 filter 804. The low pass filter may be any of the various commercial low pass
18 filters. For example, the low pass filter proposed to the Moving Picture Experts
19 Group (MPEG) for MPEG-4 may be used. The coefficients for the MPEG-4 low
20 pass filter are as follows: $c = [5/32, 11/32, 11/32, 5/32]$. Alternatively, the chroma
21 separator 800 may keep the YUV values without processing the YUV values
22 through low pass filter 804. The process for separating the base stream from
23 format A also includes a down-sampler 808. Down-sampler 808 is configured to
24 keep the chrominance blocks for each line and row specified for the desired output
25

format. The conversion of format A into base format B is known to those skilled in the art and is commonly performed today. The outcome of down-sampler 808 is the base format B stream (e.g., YUV420).

In another embodiment, filter 804 and the down-sampler 808 may also be combined into a convolution operation. In general, convolution includes a combination of multiplication, summarization, and shifting. One exemplary convolution operation is as follows:

$$L_k = c_0 * f_{2k} + c_1 * f_{2k+1} + c_2 * f_{2k+2} + c_3 * f_{2k+3} \quad \text{eq. 1}$$

Where $k=0,1,2,\dots,n-1$.

$$H_k = d_0 * f_{2k} + d_1 * f_{2k+1} + d_2 * f_{2k+2} + d_3 * f_{2k+3} \quad \text{eq. 2}$$

Where $k=0,1,2,\dots,n-1$.

At boundary pixels, mirror extension may be applied. One exemplary method for applying mirror extension for when there is an even number of taps is as follows:

$$f_{-2}=f_1, f_{-1}=f_0, f_{2n}=f_{2n-1}, f_{2n+1}=f_{2n-2} \quad \text{eq. 3}$$

Another exemplary method for applying mirror extension for when there is an odd number of taps is as follows:

$$f_{-2}=f_2, f_{-1}=f_1, f_{2n}=f_{2n-2}, f_{2n+1}=f_{2n-3} \quad \text{eq. 4}$$

1 In equations 1-4, n is the vertical dimension of the UV signal and f_k
2 corresponds to the pixel value at position k in format A chrominance blocks. L_k
3 and H_k represent pixel values at position k of the resulting base format B and
4 enhanced format B streams.

5 The process for separating the enhanced stream from format A is now
6 described. The chroma separator 800 may include an optional high pass filter 806.
7 An exemplary high pass filter 806 may have the following coefficients: $d = [5/12,$
8 $11/12, -11/12, -5/12]$. Alternatively, the chroma separator 800 may keep the YUV
9 values from the first video encoded format without applying filter 806. The
10 process for separating the enhanced stream from format A includes a down-
11 sampler 810. In one embodiment, down-sampler 810 is configured to keep all the
12 lines which down-sampler 808 did not keep. For example, when converting
13 YUV424 to YUV420, down-sampler 810 may keep all the even lines of the output
14 of the high pass filter. In the past, during the transcoding process, these “extra”
15 chrominance blocks were simply discarded. However, in accordance with the
16 present color space coding framework, these “extra” chrominance blocks become
17 the enhanced format B stream. As will be described in detail below, by
18 maintaining these “extra” chrominance blocks in a separate stream, the inefficient
19 transcoding process may be avoided when converting between two formats.

20 In another embodiment, the filter 806 and the down sampler 810 may be
21 combined into a convolution operation similar to the convolution operation
22 described above with equations 1-4 and the corresponding text.

23 In another exemplary embodiment, a wavelet transform (i.e., decomposition
24 and down sampling) may be applied that will generate the two desired output
25 formats: base format B and enhanced format B. For example, a modified 9/7

1 Daubechies wavelet transform may be applied. Additional information describing
2 the 9/7 wavelet may be obtained from the JPEG-2000 reference. The standard 9/7
3 Daubechies wavelet transform (i.e., filtering plus down-sampling) converts Format
4 A to Format B and Enhanced Format B. The low pass analysis filter coefficients
5 and high pass analysis filter coefficients are:

6
7 L (9):

8 0.026748757411,
9 -0.016864118443,
10 -0.078223266529,
11 0.266864118443,
12 0.602949018236,
13 0.266864118443,
14 -0.078223266529,
15 -0.016864118443,
16 0.026748757411

17
18 H (7):

19 0.045635881557,
20 -0.028771763114,
21 -0.295635881557,
22 0.557543526229,
23 -0.295635881557,
24 -0.028771763114,
25 0.045635881557.

19 To ensure a minimal precision loss during the transform, an integer lifting
20 scheme is used to achieve 9/7 wavelet transform. The integer lifting scheme takes
21 every intermediate result during the process and converts the results to an integer
22 either by rounding, ceiling, flooring, or clipping. An exemplary integer lifting
23 structure 1500 is illustrated in FIGURE 15. Processing is performed from left to
24 right. In FIGURE 15, dots $x_0 \sim x_9$ represent the original pixels of Format A. Dots
25 $l_0 \sim l_4$ represent pixels in Format B. Dots $h_0 \sim h_4$ represent pixels in Enhanced

1 Format B. A curved arrow represents a mirror extension. A directional branch
2 with a symbol (alpha, beta, etc) represents the application of a multiplication
3 operation with a first multiplier being a coefficient associated with the applicable
4 symbol and a second multiplier being the value of the node it leaves. A horizontal
5 branch represents the application of a carry operation for the value of one node to
6 the next stage without scaling. Branches merging at one node means all the values
7 carried in these branches are summed together to generate the value of the merging
8 node. A modification to the value k may be applied to ensure that the resulting
9 coefficients of Format B are in the range of $[0, 255]$.

10 The outcome of chroma separator 800 when Format A corresponds to
11 YUV422 and the base format corresponds to YUV420 is illustrated in FIGURE
12 10. FIGURE 10 illustrates array 200 that has been sampled in accordance with a
13 first video encoded format (e.g., video encoded format YUV422) as illustrated in
14 FIGURE 2. Each macroblock (e.g., macroblock MB_1) includes four luminance
15 blocks and two chrominance blocks: one for U and one for V. The memory layout
16 for one macroblock in format YUV422 entails four luminance blocks and four
17 chrominance blocks: $Y_1 Y_2 Y_3 Y_4 U_1 V_1 U_2 V_2$. If this YUV422 format needs
18 to be utilized by a electronic device that accepts YUV420 format (illustrated in
19 FIGURE 3), in the past, the YUV422 format was input into a transcoder that
20 decoded each chromo block, manipulated the chromo blocks, and then encoded
21 the chromo blocks again.

22 However, using the present color space coding framework, the YUV422 is
23 encoded in a new manner, graphically depicted in array 10000 as format B, which
24 includes base B and enhanced B. In contrast to prior conversion methods that
25 discarded chrominance blocks that were not needed, the present color space

1 coding framework rearranges the chrominance blocks such that the output has
2 essentially two or more streams. The first stream includes the chrominance blocks
3 for a base format, such as YUV420, generated within the chromo separator 800 via
4 the optional low pass filter 804 and the down-sampler 806. The second stream
5 includes the extra chrominance blocks from the input format, but which are not
6 used by the base format. Thus, the first stream comprises a full set of chrominance
7 blocks associated with the base format to ensure that the base format is fully self-
8 contained. The second stream is generated within the chromo separator 800 via
9 the optional high pass filter 806 and the down-sampler 810. Thus, the second
10 stream represents an enhanced stream, which, together with the first stream,
11 reconstructs the input stream (format A). As graphically depicted, the creation of
12 the base stream and the enhanced stream may occur by shuffling the chrominance
13 blocks (pixels), which manipulate the layout of the chrominance components.

14 FIGURE 9 is a block diagram of a chroma compositor for merging the base
15 format B stream and the enhanced format B stream into the first video encoded
16 format (e.g., format A). The chroma compositor 900 includes an up-sampler 904
17 and an optional synthesis filter 908 for processing the base format B stream that is
18 input into the chroma compositor 900. In addition, chroma compositor 900
19 includes an up-sampler 906 and an optional synthesis filter 910 for processing the
20 enhanced format B stream that is input into the chroma compositor 900. The
21 chroma compositor 900 also includes a merger 912 that merges the output after
22 up-sampling and filtering into the desired first video encoded format. In one
23 exemplary embodiment involving YUV424 and YUV420 formats, merger 912
24 sums up the output of two synthesis filters to re-constructs the YUV424 video
25 stream.

Up-sampler 904 pads the incoming stream as needed. The optional synthesis filter 908 may employ coefficients as follows: $c' = [-5/12, 11/12, 11/22, -5/12]$.

Up-sampler 906 also pads its incoming stream as need. The optional synthesis filter 910 may employ coefficients as follows: $d' = [-5/32, 11/32, -11/32, 5/32]$. The up-sampler 904 and the synthesis filter 908 may be merged into a convolution operations as follows:

$$f_{2k} = 2*(c_0' * L_k + c_2' * L_{k-1} + d_0' * H_k + d_2' * H_{k-1}) \quad \text{eq. 5}$$

Where $k=0,1,2,\dots,n-1$.

$$f_{2k+1} = 2*(c_1' * L_k + c_3' * L_{k-1} + d_1' * H_k + d_3' * H_{k-1}) \quad \text{eq. 6}$$

Where $k=0,1,2,\dots,n-1$.

Up-sampler 904 and 906 performs exactly the reverse operation of the down-sampler 806 and 810 respectively. For those lines discarded in 806 and 810, 904 and 906 will fill zero. After the up-sampler, the signal is restored to the original resolution.

At boundary pixels, mirror extension may be applied. One exemplary method for applying mirror extension for when there is an even number of taps, is as follows:

$$L_{-1} = L_0, H_{-1} = H_0 \quad \text{eq. 7}$$

Another exemplary method for applying mirror extension for when there is an odd number of taps, is as follows:

$$L_{-1} = L_1, H_{-1} = H_1 \quad \text{eq. 8}$$

In equations 5-8, n is the vertical dimension of the UV signal and f_k corresponds to the pixel value at position k of Format A chrominance. L_k and H_k represent pixel values at position k of the resulting base format B and enhanced format B streams.

In another embodiment for decoder 1200, an inverse 9/7 wavelet transform (i.e., up-sampling and filtering) is performed to reconstruct Format A video from the base Format B and the Enhanced Format B. The low pass synthesis filter coefficients and high pass synthesis filter coefficients are as follows:

L (7):

-0.045635881557,
-0.028771763114,
0.295635881557,
0.557543526229,
0.295635881557,
-0.028771763114,
-0.045635881557

H (9):

0.026748757411,
0.016864118443,
-0.078223266529,
-0.266864118443,
0.602949018236,
-0.266864118443,
-0.078223266529,
0.016864118443,
0.026748757411.

FIGURE 16 illustrates the corresponding integer lifting structure 1600 associated with the inverse modified 9/7 Daubechies wavelet transform. The symbols as defined for FIGURE 15 describe integer lifting structure 1600.

1 The encoder 1100 and decoder 1200 may be implemented using various
2 wavelet transforms. For example, a modified 5/3 Daubechies wavelet transform
3 may be used. FIGURES 17-18 illustrate the integer lifting structures 1700 and
4 1800 associated with the modified 5/3 Daubechies wavelet transform and the
5 inverse modified 5/3 Daubechies wavelet transform, respectively. Again, the
6 symbols as defined for FIGURE 15 describe integer lifting structures 1700 and
7 1800.

8 The corresponding low pass analysis filter coefficients and high pass
9 analysis filter coefficients are:

10 $L(5): -1/8, 1/4, 3/4, 1/4, -1/8$

11 $H(3): -1/4, 1/2, -1/4.$

12
13 The low pass synthesis filter coefficients and high pass synthesis filter
14 coefficients are:

15
16 $L(3): 1/4, 1/2, 1/4$

17 $H(5): -1/8, -1/4, 3/4, -1/4, -1/8.$

18
19 In another exemplary implementation, a 7/5 wavelet transform may be
20 used. FIGURES 19-20 illustrate the integer lifting structures 1900 and 2000
21 associated with the 7/5 wavelet transform and the inverse 7/5 wavelet transform,
22 respectively. Again, the symbols as defined for FIGURE 15 describe integer
23 lifting structures 1900 and 2000.
24
25

1 The corresponding low pass analysis filter coefficients and high pass
2 analysis filter coefficients are:

3
4
5 L(7):

6 0.0012745098039216
7 0.0024509803921569,
8 0.2487254901960785,
9 0.4950980392156863,
10 0.2487254901960785,
11 0.0024509803921569,
12 0.0012745098039216

13
14 H(5):

15 -0.13000000000000000,
16 -0.25000000000000000,
17 0.76000000000000000,
18 -0.25000000000000000,
19 -0.13000000000000000.

20
21 The low pass synthesis filter coefficients and high pass synthesis filter
22 coefficients are as follows:

23
24 L(5):

25 -0.13000000000000000,
0.25000000000000000,
0.76000000000000000,
0.25000000000000000,
-0.13000000000000000

26
27 H(7):

28 -0.0012745098039216,
29 0.0024509803921569,
30 -0.2487254901960785,
31 0.4950980392156863,
32 -0.2487254901960785,
33 0.0024509803921569,

FIGURE 11 is a block diagram of an encoder 1100 which operates in accordance with the present color space coding framework. The encoder 1100 includes a base format encoder (represented generally within box 1120), an enhanced format encoder (represented generally within box 1140), and an output bit stream formulator 1160. In addition, encoder 1100 may include a chroma separator 800 as shown in FIGURE 8 and described above. The encoder 1100 is a computing device, such as shown in FIGURE 7, which implements the functionality of the base format encoder, the enhanced format encoder, the bit stream formulator, and the optional chroma separator 800 in hardware, software or in any combination of hardware/software in a manner that produces the desired bit streams that are input into an associated decoder shown in FIGURE 12 and described below.

In overview, encoder 1100 processes two streams, the base stream and the enhanced stream, in accordance with the present color space coding framework. One advantage of encoder 1100 is the ability to provide an additional prediction coding mode, spatial prediction (SP), along with the Intra and Inter prediction coding modes. As will be described in detail below, encoder 1100 provides the spatial prediction for the enhanced chrominance blocks using the base chrominance blocks from the same frame. Due to the high correlation between the enhanced chrominance blocks and the base chrominance blocks, the spatial prediction (SP) can provide a very efficient prediction mode.

In one embodiment, encoder 1100 accepts the output streams generated from the chroma separator 800. In another embodiment, chroma separator 800 is included within encoder 1100. For either embodiment, chroma separator 800

1 accepts input encoded in a first encoded format 1106, referred to as format A. The
2 generation of the first encoded format 1106 is performed in a conventional manner
3 known to those skilled in the art of video encoding. In certain situations, the
4 generation of the first encoded format is accomplished by converting a format
5 from another color space, such as the RGB color space. When this occurs, a color
6 space converter (CSC) 1104 is used. The color space converter 1104 accepts an
7 input 1102 (e.g., RGB input) associated with the other color space. The color
8 space converter 1104 then converts the input 1102 into the desired first encoded
9 format 1106. The color space converter 1104 may use any conventional
10 mechanism for converting from one color space to another color space. For
11 example, when the conversion is between the RGB color space and the YUV color
12 space, the color space converter 1104 may apply known transforms that are often
13 represented as a set of three equations or by a matrix. One known set of equations
14 defined by one of the standards is as follows:

$$15 \quad Y = 0.299 \times R + 0.587 \times G + 0.114 \times B$$

$$16 \quad U = -0.299 \times R - 0.587 \times G + 0.886 \times B$$

$$17 \quad V = 0.701 \times R - 0.587 \times G - 0.114 \times B.$$

18
19 The transform is also reversible, such that given a set of YUV values, a set
20 of RGB values may be obtained. When a color space conversion is necessary, the
21 processing performed by the chroma separator 800 may be combined with the
22 processing performed in the color space converter 1104. The chroma separator
23 800 and color space conversion 1804 may be included as elements with encoder
24
25

1 1100. Alternatively, encoder 1100 may accept the outputs generated by the
2 chroma separator 800.

3 As described above in conjunction with FIGURE 8, the chroma separator
4 800 is configured to output a base format stream 1108 and at least one enhanced
5 format stream 1110. The base format stream 1108 is processed through the base
6 encoder 1120 and the enhanced format stream is processed through the enhanced
7 encoder 1140.

8 Base encoder 1120 is any conventional encoder for the base format stream
9 1108. In general, base encoder 1120 attempts to minimize the amount of data that
10 is output as the base bit stream (B-BS), which will typically be transmitted through
11 some media so that the encoded video may be played. The conventional base
12 encoder 1120 includes conventional elements, such as a discrete cosine transform
13 (DCT) 1122, a quantization (Q) process 1124, a variable length coding (VLC)
14 process 1126, an inverse quantization (Q^{-1}) process 1128, an inverse DCT (IDCT)
15 1130, a frame buffer 1132, a motion compensated prediction (MCP) process 1134,
16 and a motion estimation (ME) process 1136. While the elements of the base
17 encoder 1120 are well known, the elements will be briefly described to aid in the
18 understanding of the present color space coding framework.

19 However, before describing the conventional base encoder 1120,
20 terminology used throughout the following discussion is defined. A frame refers
21 to the lines that make up an image. An Intraframe (I-frame) refers to a frame that
22 is encoded using only information from within one frame. An Interframe, also
23 referred to as a Predicted frame (P-frame), refers to a frame that uses information
24 from more than one frame.
25

1 Base encoder 1120 accepts a frame of the base format 1108. The frame will
2 be encoded using only information from itself. Therefore, the frame is referred to
3 as an I-frame. Thus, the I-frame proceeds through the discrete cosine transform
4 1122 that converts the I-frame into DCT coefficients. These DCT coefficients are
5 input into a quantization process 1124 to form quantized DCT coefficients. The
6 quantized DCT coefficients are then input into a variable length coder (VLC) 1126
7 to generate a portion of the base bit stream (B-BS). The quantized DCT
8 coefficients are also input into an inverse quantization process 1128 and an inverse
9 DCT 1130. The result is stored in frame buffer 1132 to serve as a reference for P-
10 frames.

11 The base encoder 1120 processes P-frames by applying the motion
12 estimation (ME) process 1134 to the results stored in the frame buffer 1132. The
13 motion estimation process 1134 is configured to locate a temporal prediction (TP),
14 which is referred to as the motion compensated prediction (MCP) 1134. The MCP
15 1134 is compared to the I-frame and the difference (i.e., the residual) proceeds
16 through the same process as the I-frame. The motion compensated prediction
17 (MCP) 1134 in the form of a motion vector (MV) is input into the variable length
18 coder (VLC) 1126 and generates another portion of the base bit stream (B-BS).
19 Finally, the inverse quantized difference data is added to the MCP 1134 to form
20 the reconstructed frame. The frame buffer is updated with the reconstructed
21 frame, which serves as the reference for the next P-frame. It is important to note
22 that the resulting base bit stream (B-BS) is fully syntactically compatible with
23 conventional decoders available in existing devices today that decode base stream
24 B format.

Enhanced encoder 1140 attempts to minimize the amount of data that is output as the enhanced bit stream (E-BS). This enhanced bit stream is typically transmitted through some media, and optionally decoded, in order to play the higher quality encoded video. While having an enhanced encoder 1140 within encoder 1100 has not previously been envisioned, enhanced encoder 1140 includes several conventional elements that operate in the same manner as described above for the base encoder. The conventional elements include as a discrete cosine transform (DCT) 1142, a quantization (Q) process 1144, a variable length coding (VLC) process 1146, an inverse quantization (Q^{-1}) process 1148, an inverse DCT (IDCT) 1150, a frame buffer 1152, and a motion compensated prediction (MCP) process 1154. One will note that a motion estimation process is not included within the enhanced encoder 1140 because the enhanced stream does not include any luminance blocks containing the Y component. Motion vectors (MVs) are derived from Y components. However, in accordance with the present color space coding framework, enhanced encoder 1140 includes a mode selection switch 1158 that selectively predicts a P-frame. Switch 1158 may select to predict the P-frame from a previous reference generated from the enhanced stream stored in frame buffer 1152 or may select to “spatially” predict (SP) the P-frame using a reference from the base stream that is stored in the frame buffer 1132 for the current frame. Spatial prediction provides a very efficient prediction method due to the high correlation between enhanced chrominance blocks in the enhanced stream and chrominance blocks in the base stream. Thus, the present color space coding framework provides greater efficiency in prediction coding and results in a performance boost in comparison to traditional encoding mechanisms. The output of enhanced encoder 1140 is the enhanced bit stream (E-BS).

1 Although the conventional elements in the base encoder 1120 and the
2 enhanced encoder 11140 are illustrated separately, in one embodiment, the base
3 encoder 1120 and the enhanced encoder 1140 may share one or more of the same
4 conventional elements. For example, instead of having two DCTs 1122 and 1142,
5 one DCT may be used by both the base encoder 1120 and by the enhanced encoder
6 1140. Thus, developing an encoder 1100 in accordance with the present color
7 space coding framework requires minimal extra effort in either hardware,
8 software, or any combination to accommodate the enhanced stream. In addition,
9 other advanced encoding techniques developed for the base encoder 1220 can be
10 easily applied to the present color space coding framework. For example, the
11 present color space coding framework operates when there are bi-directionally
12 predicted frames (B-frames).

13 The output bit stream formulator 1160 combines the enhanced bit stream
14 (E-BS) with the base bit stream (B-BS) to form a final output bit stream.
15 Exemplary formats for the final output bit stream are illustrated in FIGURES 13
16 and 14 and are described in conjunction with those figures.

17 FIGURE 12 is a block diagram of a decoder which incorporates the present
18 color space coding framework. In overview, the decoder 1200 may perform a
19 simple bit stream truncation to obtain the lower quality video format. Thus, the
20 expensive transcoding process is not necessary. In general, decoder 1200 reverses
21 the process performed by encoder 1100. Decoder 1200 accepts the base bit stream
22 (B-BS) and the enhanced bit stream (E-BS). The base bit stream and the enhanced
23 bit stream may have been parsed with an input bit stream parser 1202 included
24 within the decoder or external to the decoder. The decoder 1200 includes a base
25 format decoder (represented generally within box 1220) and an enhanced format

1 decoder (represented generally within box 1240). The base decoder 1220
2 processes the base bit stream and the enhanced decoder 1240 processes the
3 enhanced bit stream. In addition, decoder 1200 may include a chroma compositor
4 900 as shown in FIGURE 9 and described above. The decoder 1200 is a
5 computing device, such as shown in FIGURE 7, which implements the
6 functionality of the base format decoder, the enhanced format decoder, and the
7 optional chroma compositor 900 in hardware, software or in any combination of
8 hardware/software in a manner that produces the desired format A 1260.

9 In overview, decoder 1200 inputs two streams, the base bit stream (B-BS)
10 and the enhanced bit stream (E-BS) generated in accordance with the present color
11 space coding framework. The decoder 1200 has the ability to decode the
12 prediction coding mode, spatial prediction (SP), provided by the encoder 1100.

13 In one embodiment, decoder 1200 includes the chroma compositor 900. In
14 another embodiment, the chroma compositor 900 is a separate device from the
15 decoder 1200. For either embodiment, chroma compositor 900 accepts the two
16 streams containing the values for the luminance blocks and chrominance blocks
17 for a base format and the values for the chrominance blocks for the enhanced
18 format and merges them into format A 1260 as explained in conjunction with
19 FIGURE 9. In certain situations, format A 1260 is converted into a format of
20 another color space, such as the RGB color space. When this occurs, a color space
21 converter (CSC) 1262 is used. The color space converter 1262 accepts format A
22 1260 as an input and converts input 1260 into output 1264 (e.g., RGB output),
23 which is associated with the other color space. The color space converter 1262
24 may use any conventional mechanism for converting from one color space to
25 another color space. For example, when the conversion is between the RGB color

1 space and the YUV color space, the color space converter 1262 may apply known
2 transforms as described above. When a color space conversion is necessary, the
3 processing performed by the chroma compositor 900 may be combined with the
4 processing performed in the color space converter 1262. The chroma compositor
5 900 and color space conversion 1262 may be included as elements within decoder
6 1200. Alternatively, decoder 1200 may supply inputs to an external the chroma
7 compositor 900.

8 Base decoder 1220 is any conventional encoder for the base bit stream (B-
9 BS). In general, base decoder 1220 reconstructs the YUV values that were
10 encoded by the base encoder 1120. The conventional base decoder 1220 includes
11 conventional elements, such as a variable length decoding (VLD) process 1222, an
12 inverse quantization (Q^{-1}) process 1224, an inverse discrete cosine transform
13 (IDCT) 1226, a frame buffer 1228, and a motion compensated prediction (MCP)
14 process 1230. Again, the elements of the base decoder 1220 are well known.
15 Therefore, the elements will be briefly described to aid in the understanding of the
16 present color space coding framework.

17 The base decoder 1220 inputs the base bit stream into the variable length
18 decoder (VLD) 1222 to retrieve the motion vectors (MV) and the quantized DCT
19 coefficients. The quantized DCT coefficient are input into the inverse quantization
20 process 1224 and the inverse DCT 1226 to form the difference data. The
21 difference data is added to its motion compensated prediction 1230 to form the
22 reconstructed base stream that is input into the chromo compositor 900. The result
23 is also stored in the frame buffer 1228 to server as a reference for decoding P-
24 frames.
25

Enhanced decoder 1240 reconstructs the UV values that were encoded by the enhanced encoder 1140. While having an enhanced decoder 1240 within decoder 1200 has not been previously envisioned, enhanced decoder 1240 includes several conventional elements that operate in the same manner as described above for the base decoder 1220. The enhanced decoder 1240 includes conventional elements, such as a variable length decoding (VLD) process 1242, an inverse quantization (Q^{-1}) process 1244, an inverse discrete cosine transform (DCT) 1246, a frame buffer 1248, and a motion compensated prediction (MCP) process 1250.

The flow of the enhanced bit stream through the enhanced decoder 1240 is identical to the base decoder 1220, except that the difference data may be selectively added to its motion compensated prediction (MCP) or added to its spatial prediction (SP), as determined by the mode information switch 1252. The outcome of the enhanced decoder 1240 is the reconstructed enhanced stream that contains the values for the “extra” chrominance blocks for the current frame.

The base stream and the enhanced stream are then input into the chroma compositor, which processes the streams as described above to reconstruct format A. Although the conventional elements in the base decoder 1220 and the enhanced decoder 1240 are illustrated separately, in one embodiment, the base decoder 1220 and the enhanced decoder 1240 may share one or more of the same conventional elements. For example, instead of having two inverse DCTs 1226 and 1246, one inverse DCT may be used by both the base decoder 1220 and by the enhanced decoder 1240. Thus, developing a decoder in accordance with the present color space coding framework requires minimal extra effort in either hardware, software, or any combination to accommodate the enhanced stream. In

1 addition, other advanced decoding techniques developed for the base decoder
2 1420 can be easily applied to the present color space coding framework. For
3 example, the present color space coding framework operates when there are bi-
4 directionally predicted frames (B-frames).

5 Thus, by coding formats using the present color space coding framework,
6 the conversion between two formats may be achieved via bit truncation, rather
7 than the expensive transcoding process. Thus, there is no transcoding process
8 performed on the formats to convert from one to another.

9 Exemplary Bit Streams

10 It is envisioned that the output bit stream formation process 1160 shown in
11 FIGURE 11 may organize the resulting base bit stream (B-BS) and the enhanced
12 bit stream (E-BS) in numerous ways. FIGURES 13 and 14 illustrate two
13 exemplary bit streams. For convenience, the exemplary bit streams illustrate the
14 organization of the base bit stream, in relation to the enhanced bit stream, and omit
15 other information that is commonly included in transport stream packets, such as
16 packet identifiers, sequence numbers, and the like. In addition, exemplary bit
17 streams may include an indicator that indicates that the bit stream supports format
18 A and base format B.

19 FIGURE 13 is a graphical representation of an exemplary bit stream 1300
20 for transmitting the multiple bit streams shown in FIGURES 11 and 12. In
21 overview, bit stream 1300 embeds the enhanced bit stream (E-BS) within the base
22 bit stream (B-BS). Thus, bit stream 1300 includes B-BS information 1302, 1304,
23 and 1306, which alternates with E-BS information 1312, 1314, and 1316. In
24 practice, if the base bit stream corresponds to YUV420 format and the enhanced
25 bit stream includes chrominance blocks for YUV422 format, bit stream 1300

1 allows a YUV422 decoder to sequentially decode all the frames. However, a
2 YUV420 decoder that decodes bit stream 1300 must skip the E-BS frames. Bit
3 stream 1300 is suitable for streaming/broadcasting applications.

4 FIGURE 14 is a graphical representation of another exemplary bit stream
5 1400 for transmitting the multiple bit streams shown in FIGURES 11 and 12. In
6 overview bit stream 1400 concatenates the enhanced bit stream to the end of the
7 base bit stream. Thus bit stream 1400 includes consecutive frames of base bit
8 stream (e.g., frames 1402, 1404, 1406) followed by consecutive frames of
9 enhanced bit stream (e.g., frames 1412, 1414, 1416). In practice, if the base bit
10 stream corresponds to the YUV420 format and the enhanced bit stream includes
11 chrominance blocks for the YUV422 format, bit stream 1400 allows a YUV420
12 decoder to sequentially decode all the frames without encountering the enhanced
13 bit stream. The YUV420 can terminate the decoding process after all the base bit
14 frames (e.g., 1402, 1404, and 1406) are decoded. However, a YUV422 decoder
15 must seek and decode the base bit stream and the enhanced bit stream before
16 proceeding to the next frame. The YUV422 decoder may utilize two pointers to
17 sequentially access the base bit stream and the enhanced bit stream. Bit stream
18 1400 is suitable for down-and-play applications.

19 Bit stream 1400 may also be separated into different individual files. In
20 this embodiment, the base bit stream represents a standalone stream and would be
21 fully decodable by a YUV420 decoder and would not require any modifications to
22 existing YUV420 decoders. A YUV422 decoder would process the two bit stream
23 files simultaneously. Bit stream 1400 may be advantageously implemented within
24 video recording devices, such as digital video camcorders. Bit stream 1400 would
25 allow recording both a high quality and low quality stream. If a consumer realizes

1 that additional recording is desirable but the current media has been consumed, an
2 option on the digital video camcorder may allow the consumer to conveniently
3 delete the high quality stream and keep the low quality stream so that additional
4 recording may resume.

5 The following description sets forth a specific embodiment of a color space
6 coding framework that incorporates elements recited in the appended claims. The
7 embodiment is described with specificity in order to meet statutory requirements.
8 However, the description itself is not intended to limit the scope of this
9 patent. Rather, the inventors have contemplated that the claimed invention might
10 also be embodied in other ways, to include different elements or combinations of
11 elements similar to the ones described in this document, in conjunction with other
12 present or future technologies.